Information Fusion and Extraction Priorities for Australia's Information Capability

Martin Oxenham[1], John Percival[1], Richard Price[2] and Dale Lambert[2]
[1]Intelligence, Surveillance & Reconnaissance Division
[2]Command & Control Division
Defence Science and Technology Organisation
West Avenue, Edinburgh
South Australia 5111
AUSTRALIA

{martin.oxenham, john.percival, richard.price, dale.lambert}@dsto.defence.gov.au

ABSTRACT

The Commonwealth of Australia's strategic policy on defence recognises the need for Australia to further develop and enhance its information capability. While this is mainly a reflection of the uptake of information technology by the military (the Revolution in Military Affairs), in more recent times it has also been in response to the global threats of terrorism and the utilisation of weapons of mass destruction or long-range ballistic missiles by rogue states. The areas which have been identified as being of primary importance for fostering Australia's emerging information capability are intelligence and surveillance capabilities, communications, information warfare, command and headquarters systems, and logistics and business applications. To meet the challenges that the development and enhancement of this information capability raise, advanced information processing techniques for fusing and extracting data or information are required. In this paper, a holistic model for integrating the relevant technologies of data fusion and data mining is proposed and several of the current information fusion and extraction initiatives at Australia's Defence Science and Technology Organisation supporting intelligence, surveillance, and command and control are outlined.

1.0 INTRODUCTION

Australia's strategic policy, as laid down in its Defence White Paper [1], identifies the need for Australia to further develop and enhance its information capability. It states that "Effective use of information is at the heart of Australia's defence capability. In part this is a reflection of a worldwide trend, as information technology is transforming the ways in which armed forces operate at every level. All forms of capability are being transformed by the innovative use of information technology. But this trend is more significant to Australia than to many other countries. Our strategic circumstances mean that innovative applications of different aspects of information technology offer Australia unique advantages." [1, p. 94, para. 8.78]

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
00 MAR 2004	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Information Fusion and Extraction Priorities for Australias Information Capability	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Defence Science and Technology Organisation West Avenue, Edinburgh South Australia 5111 AUSTRALIA**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001673, RTO-MP-IST-040, Military Data and Information Fusion (La fusion des informations et de données militaires)., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**41**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

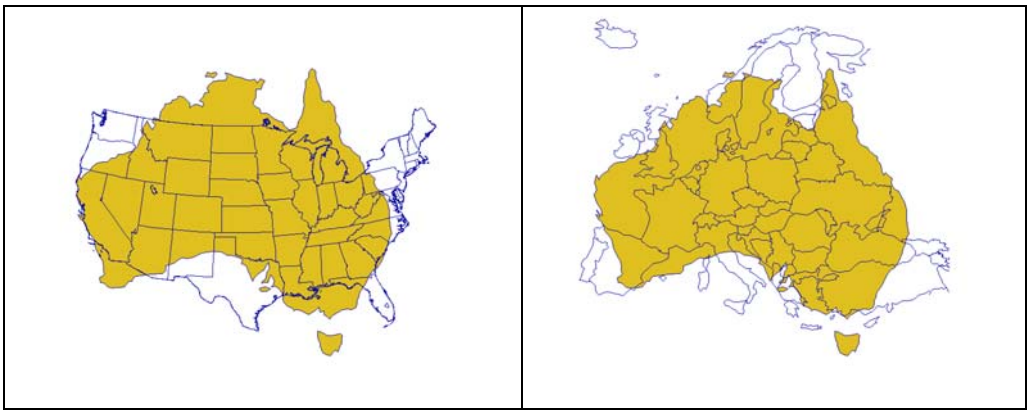


Figure 1: Geographic Comparison of Australia to the USA, the UK and Europe[1].

So what are Australia's strategic circumstances? In terms of geography, Australia is an island continent that ranks sixth in size on a worldwide scale. It covers an area comparable to both that of the USA and of Europe, as depicted in Fig. 1. However, in terms of population, Australia may be regarded as a small nation. Australia has a current population of only 19.9 million, which is 14 times less than that of the USA, 30 times less than that of the whole of Europe, and between 2 and 4 times less than that of the UK, and some of Europe's larger nations including France, Germany, Italy and Spain. The percentage of Australia's population in the permanent defence force is comparable to the percentages for the other countries listed above, but when the defence reserves are factored in, Australia is way below par (refer to Table 1). In terms of actual numbers, Australia's combined defence force is substantially smaller than the average crowd at some premier Australian sporting events such as the Australian Football League's grand final which typically attracts close to 100,000 spectators.

Table 1: Australia compared to other nations based on the sizes of their populations and defence forces in July 2002 [3].

Nation	Population	Size of the Permanent Defence Force	Size of the Defence Force Reserves	% of the Population in the Permanent Defence Force	% of the Population in the Combined Defence Force
Australia	19,546,792	55,200	27,730	0.3	0.4
USA	280,562,489	1,371,500	1,303,300	0.5	1.0
UK	59,778,002	212,400	191,000	0.4	0.7
France	59,765,983	317,300	419,000	0.5	1.2
Germany	83,251,851	332,800	344,000	0.4	0.8
Italy	57,715,625	265,000	72,000	0.5	0.6
Spain	40,077,100	186,000	447,900	0.5	1.6

The majority of the Australian population resides on the eastern seaboard, which is also problematic from a defence perspective, since the most vulnerable stretch of Australia's vast coastline and territorial waters is that which lies to the north. This presents major challenges to the Australian Defence Force and the various Australian intelligence agencies which jointly serve to protect Australia's sovereign territory and national interests. Current defence imperatives for the nation include border protection against smuggling (of people, drugs and other contraband) and illegal fishing, peacekeeping efforts such as in East Timor and increasingly homeland defence in the wake of terrorist attacks by extremist groups, and the threat of the

[1] The images have been produced using Albers' equal-area conic projection [2, p.49].

utilisation of long-range ballistic missiles and weapons of mass destruction by rogue states [4]. Accordingly, "maintaining first-rate intelligence capabilities, developing a comprehensive surveillance system providing continuous coverage of our extended air and sea approaches, developing an integrated command system covering operations at all levels and in all environments, and maximising the efficiency of our logistics systems and management processes by cost-effective investment in information technology applications" [1, p. 95, para. 8.83] are specific goals of Australia's emerging information capability. To achieve these goals, innovative methods for automatically extracting and fusing information from voluminous, disparate and distributed data sources are required to assist defence force personnel and intelligence analysts to perform their duties more efficiently and effectively. The purpose of this paper is to outline some of the initiatives at Australia's Defence Science and Technology Organisation (DSTO)[2] which are striving to meet these goals.

The remainder of the paper is structured as follows. In Section 2, the concepts and terminology used in the paper are presented, and a holistic model for integrating data fusion and data mining is proposed. In Sections 3-5, several of the pertinent information fusion and extraction initiatives at the DSTO in the areas of wide area surveillance, intelligence analysis, and command and control are described. Finally, in Section 6, some concluding remarks are made.

2.0 DATA FUSION AND DATA MINING

The purpose of an information capability for the defence and intelligence communities is to support the establishment and enhancement of the operators' and analysts' situation awareness, and ultimately to facilitate their or others' timely and effective decision making. To assist in the processing of the large volumes of data and information required for meeting these goals, the key enabling technologies of data fusion and data mining have emerged over the last decade or so.

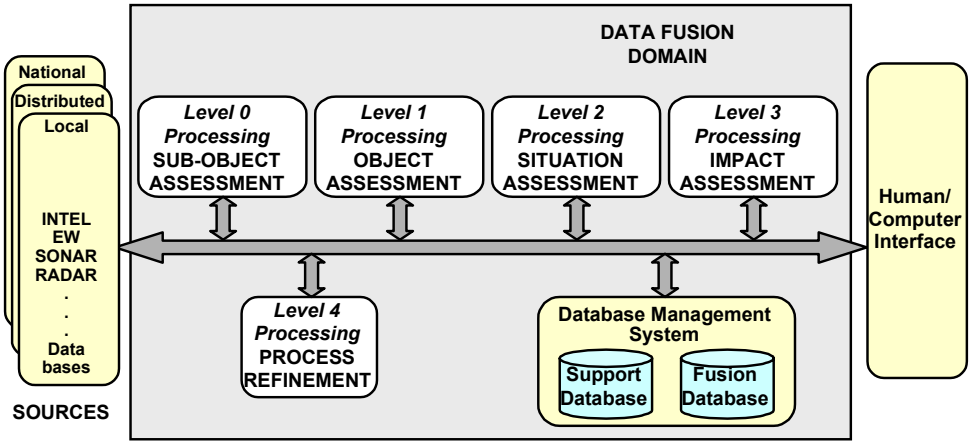


Figure 2: The United States Joint Director of Laboratories (JDL) Model of Data Fusion [8].

Data fusion is commonly described as the process of combining data or information to estimate or predict entity states [5, 6]. The functional model of data fusion introduced by the United States Joint Directors of Laboratories in 1987 [7], and revised in 1998 [8], has gained widespread acceptance within the data fusion community. It represents data fusion in terms of the five functional levels of *signal refinement* (Level 0), *object refinement* (Level 1), *situation refinement* (Level 2), *impact refinement* (Level 3) and *process*

[2] The DSTO constitutes part of the civilian sector of the Australian Defence Organisation. Its role within Defence is to ensure the expert, impartial and innovative application of science and technology to the defence of Australia and its national interests. It accomplishes this via a combination of means including prototype capability development, project support for the acquisition of defence materiel, and long range research.

refinement (Level 4) (refer to Fig. 2). It is noted that Level 1 is also often referred to as *sensor fusion* and Levels 2 and 3 jointly as *information fusion*.

Data mining, which subsumes machine learning and data visualisation, may be regarded as the process of determining patterns, trends, relationships and associations in large data sets that are not explicitly mentioned in the raw data. Thus, in contrast to data fusion as defined above, which is in broad terms a "supervised" process relying on models (eg target motion models, ontologies, inference rules, plans), data mining is fundamentally an "unsupervised" process which discovers models. Furthermore, unlike data fusion, no model of data mining has been developed to the knowledge of the authors.

Despite the distinction between data fusion and data mining that these definitions lead to, there is a strong interdependence between the two processes. Little research appears to have focussed on their integration however. One exception to this is the integrated model that Waltz proposes in [9], which is in keeping with the principles of knowledge discovery in databases (KDD)[3] [10, 11]. In his model, depicted in Fig. 3, sensor data is fed into both the data fusion and data mining modules for separate processing. While the data are analysed by the data fusion module to support decision making in the application domain, they are also passed to the data mining module for data warehousing. Selected data from the warehouse are then mined to hypothesise models that classify the entities represented by the data or the relationships between those entities; these hypothesised models then pass through a validation stage, and if validated are then evaluated and interpreted. Finally, the discovered models are passed to Level 0, 1 or 2 of the fusion model as appropriate to enhance the overall data fusion process.

While this is a fair model of integrated data fusion and data mining, it has several shortcomings. First, it is not apparent why it does not entertain the possibility of discovering models that may feed into Level 3 of the data fusion process. Second, it fails to allow established models in the data fusion process to assist in the data mining process. Finally, it has not attempted to view the integration of data fusion and data mining holistically, but instead has simply coupled the two processes together. In the remainder of this section, a strategic framework is outlined for dealing with the integration problem from a holistic perspective.

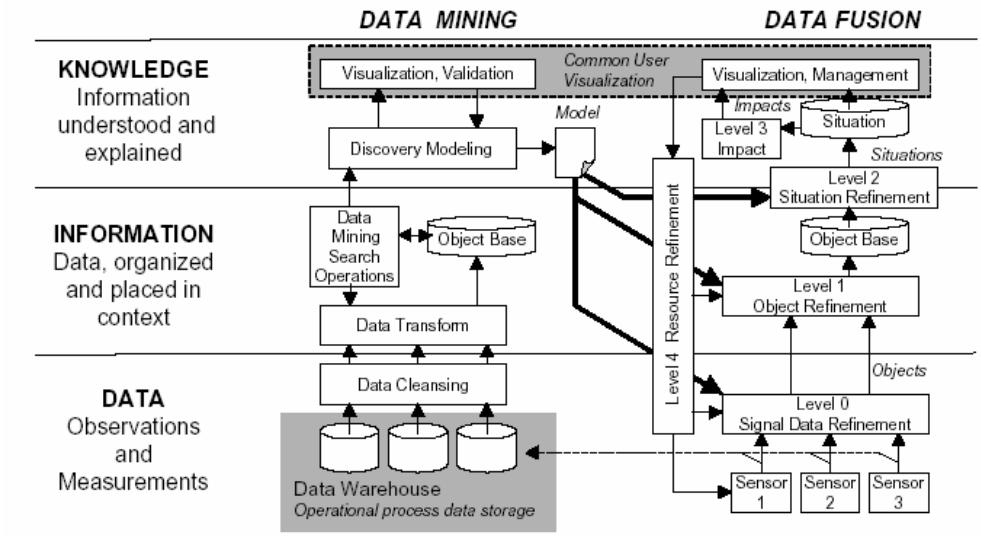


Figure 3: Waltz' Integrated Model of Data Mining and Data Fusion [9].

[3] The process of KDD involves a number of steps including data warehousing, target data selection, cleaning and pre-processing, transformation and reduction (summarisation), data mining, model selection (or combination), evaluation and interpretation, and finally consolidation and use of the extracted "knowledge" [11].

In contrast to the definition of data fusion presented above, Lambert [12] defines data fusion as the process of utilising one or more data sources over time to assemble a representation of aspects of interest in an environment. He also offers a deconstruction of the US JDL model of data fusion described earlier in the section, in which Level 0 is included in Level 1, and Level 4 is absorbed into each of Levels 1, 2 and 3 (refer to Fig. 4). Under this deconstructed JDL account, denoted hereafter by λJDL, Level 1 is about the identification of objects from their properties, Level 2 is about the identification of relations between these objects, and Level 3 is about the identification of the effects of these relationships between the objects. Articulated in more detail [12]:

- *Object refinement* is the process of utilising one or more data sources over time to assemble a representation of objects of interest in an environment. An *object assessment* is a stored representation of objects obtained through object refinement;
- *Situation refinement* is the process of utilising one or more data sources over time to assemble a representation of relations between objects of interest in an environment. A *situation assessment* is a stored representation of relations between objects obtained through situation refinement; and
- *Impact refinement* is the process of utilising one or more data sources over time to assemble a representation of effects of situations in an environment, relative to one's intentions. An *impact assessment* is a stored representation of effects of situations obtained through impact refinement.

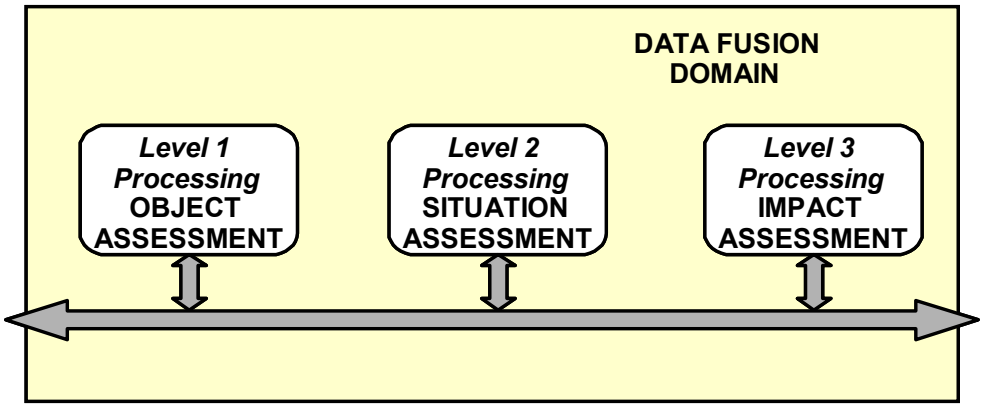


Figure 4: The λJDL Model of Data Fusion [12].

Fundamentally, the λJDL model of data fusion is not at odds with the JDL model, but proves to be far more convenient and flexible; this point will be expanded upon later in Section 5. Perhaps more interestingly, the λJDL description of the refinement processes at each level not only allows for the combination of data or information as in the JDL model, but the notion of "utilising one or more data sources over time to assemble a representation" also accommodates the use of the data or information at each level for developing or enhancing models. Thus, under the λJDL model of data fusion, data mining may be regarded as an integrated sub-process. There are a number of benefits in adopting this view. First, the data mining process acquires more structure through the inheritance of the notion of "level" from the data fusion model. Second, while the traditional data fusion and data mining processes may be distinguished within the λJDL model if required, there is no compulsion to do so. As such, more effort may be directed towards the appropriate utilisation of the data or information to meet the decision-maker's needs without having to be concerned about whether the processing is "fusion or mining". Third, analogous concepts shared by traditional data fusion and data mining such as "data alignment (data fusion)" and "data cleansing and transformation (data mining)", "multi-sensor tracking and track fusion (data fusion)" and "multi-source information retrieval (data mining)", and "information extraction (both)" for example may be either legitimately identified or classified as specific instances of general data

fusion concepts. Finally, challenges common to both processes such as knowledge representation, and data or information visualisation for example may be dealt with in unison. For all of these reasons, this is the viewpoint that is taken in this paper.

3.0 TARGET BEHAVIOUR EXTRACTION FOR WIDE AREA SURVEILLANCE

As highlighted in the introduction, wide area surveillance of the sea-air gap to the north of Australia plays a key role in Australia's information capability strategy. Wide area surveillance in the Australian context is currently afforded by a suite of surveillance assets, primarily 2- and 3-dimensional ground-based microwave radars, and three over-the-horizon radars (two of which constitute the Jindalee Operational Radar Network (JORN) that became operational several months ago); refer to Fig. 5 for a stylised depiction of the radar coverage regions. These will be augmented in the coming years by an early warning and control capability in the form of a fleet of Wedgetail AEW&C platforms, and a Link 16 capability. The wide area surveillance picture (WASP), formed through the fusion of data from these sensors and additional sources, supports a number of C[4]ISR activities[4] namely real-time surveillance, fighter control and situation and threat refinement at the tactical level, and mission planning and intelligence gathering at the theatre level. Typically, compilation of the WASP and these subsequent C[4]ISR activities are performed manually by air defence operators. To allow the operators to focus more on decision making, and less on data and information integration, it is desirable to automate some of the data and information processing required for these activities, providing it does not lead to a loss of their situation awareness. In particular, to support automated situation and threat refinement and intelligence gathering, techniques for automatically interpreting current activities and predicting future events are required. In this section, one such technique is described which takes the approach of exploiting the contextual information in the surveillance region to "extract" the behaviours of the detected air platforms.

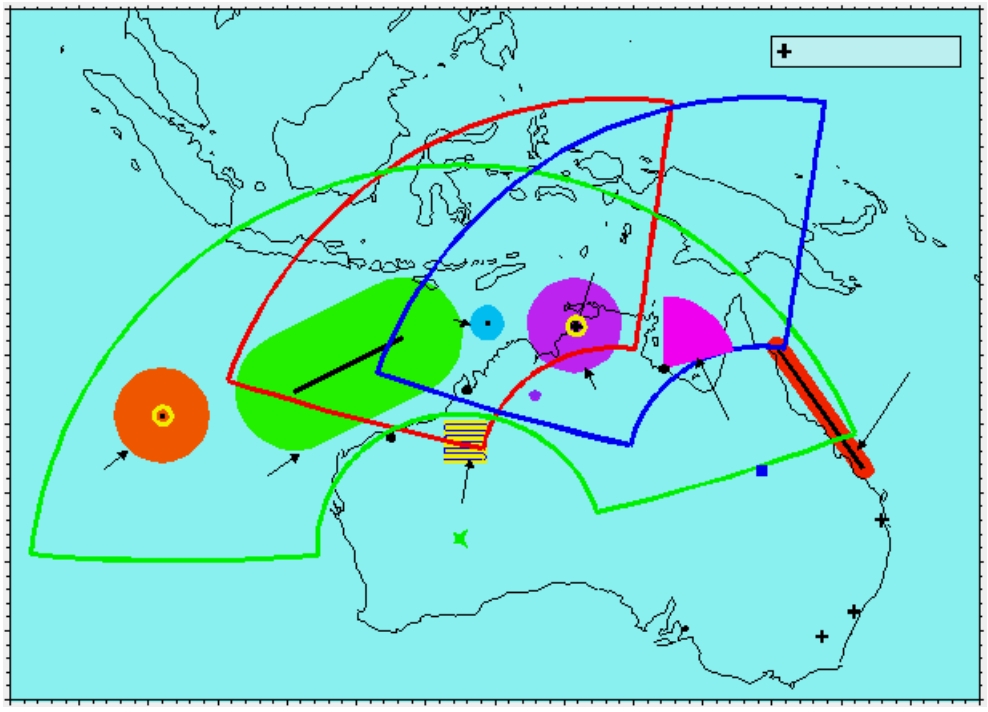


Figure 5: Stylised depiction of Australia's wide area surveillance system [16].

[4] C[4]ISR stands for Command, Control, Communications, Computers, Intelligence, Surveillance & Reconnaissance.

Before outlining the methodology, it is instructive to consider how human operators accomplish the task. When human operators compile the WASP, they do not rely solely on filtered data from radars and other sensors. The track data are plotted on a geographic display which provides the operators with a wealth of contextual information to exploit for enhancing their situation awareness. This includes entities such as coastlines, air and shipping lanes, locations of towns/cities, air and shipping ports, et cetera, locations of military assets and installations, and regions such as air defence identification zones, fighter and missile engagement zones et cetera, and notional radar coverage regions. This information enables them to visually infer relations between the target tracks and the entities (including other target tracks) in order to recognise activities and events that are occurring within the surveillance region. It may also allow them to interpret the significance of those events in terms of evidence they may provide of target identity or class, as well as the threat they may pose or the impact they may have on operations of friendly and coalition forces. Finally, it may also assist them in identifying and correcting sensor misalignments, and scheduling agile sensors or cueing additional sensors. Thus, the contextual information provides support for operators at all levels of the λJDL model of data fusion. In order to introduce automation into this processing for assisting operators, algorithms are required which are capable of extracting the same types of relations or predicates between tracks and entities that the operators obtain via visual means.

To develop appropriate means for modelling and utilising the available contextual information for extracting target-entity relations, it is necessary to determine first what types of relations require automatic recognition. While no compiled list of relations can ever be truly exhaustive, those listed in Table 2, along with their auxiliary functions, have provided a firm basis from which to work:

Table 2: Entities, relations and functions used for extracting symbolic information from sensor tracks via contextual information.

ENTITIES	RELATIONS
Points, line segments, circular arcs, circles, polygons, annular sectors, and apertures	Is_Above_Height (or Speed) Is_Below_Height (or Speed) Is_Between_Heights (or Speeds)
FUNCTIONS	
Time_To_Go Time_Elapsed Point_Of_Ingress Point_Of_Egress Minimum_Distance_To Minimum_Time_To Point_Of_Realisation	Is_In or Is_On Is_Heading_To Is_Heading_From Is_Passing Is_Within_Angle Is_Within_Distance Will_Be_Within_Distance

The issue of how to model the contextual information listed at the beginning of this section[5] is somewhat dependent on the use to which the information will be put [13-17]. However, the main approach being taken to this problem by DSTO is to model the contextual information geometrically as the composition of simple geometric entities such as points, line segments, circular arcs, apertures, and circular, annular and

[5] It is noted that additional contextual information may also be available for use by air defence operators, such as the order of battle (ORBAT) and the electronic order of battle (EOB), and information about a target's weapon and speed envelopes that may be inferred from its classification or identification.

polygonal regions [16]. For example, to a first approximation, the location of an asset may be modelled as a point, an airlane may be modelled as a rectangle (along with a specified direction), the notional coverage region of a ground-based microwave radar may be modelled as a circle, the weapons envelope of a fighter aircraft may be modelled as an annular sector, and the racetrack orbit of a force multiplier may be modelled as the boundary of the union of a rectangle and two circles. Therefore, to support the recognition of target behaviours with respect to general entities, it is sufficient to develop functions for recognising them with respect to simpler entities and then to compose the functions to recognise the more complex behaviours. The recognition of target behaviours via this methodology is referred to as *target behaviour extraction*.

The methodology is explained in detail in [16] and [18]. However, it may be summarised as follows. Assume initially that the precise target state is known. Since the contextual information can be modelled geometrically, and the target state itself is geometric by nature, it is possible to use geometric criteria to establish mathematical conditions which hold if and only if the target is exhibiting the behaviour, that is if and only if the Boolean target-entity relation holds. Thus for a given relation *Rel*, the value of *Rel* (target, entity) is 1 or TRUE if the target is exhibiting the behaviour and is 0 or FALSE otherwise. For example, the relation "is_in" between a target with position (x_1, x_2) and a circle with centre (c_1, c_2) and radius R is 1 (or TRUE) if and only if the mathematical condition $(x_1 - c_1)^2 + (x_2 - c_2)^2 \leq R^2$ holds. To establish relations for compound target behaviours, the logical AND, OR and NOT connectives can be used accordingly to compose the truth values of the Boolean relations corresponding to each of the contributing simple target behaviours. To extend this approach to handle target state estimates, the target behaviour extraction algorithms can be applied to each of a specified number of random samples distributed according to the state estimate's probability density. The proportion of samples exhibiting the behaviour to the total number of samples can then be interpreted as the probability that the target is exhibiting the behaviour, given the state estimate and the model of the behaviour. Figure 6 below illustrates the methodology applied to the problem of monitoring the progress of an air target as it approaches the air defence identification zone (ADIZ) established around the coastline of an island during an air defence scenario. As time evolves, target-entity relations are used to determine if the target is heading towards the ADIZ (and the time to go until it enters the ADIZ, assuming it continues to fly at constant speed on a constant heading), if it is inside the ADIZ or if it is within a specified distance of the island. A simple tracking algorithm is used to estimate the target's state at 10 second intervals. The results for a single run are plotted in Fig. 6.

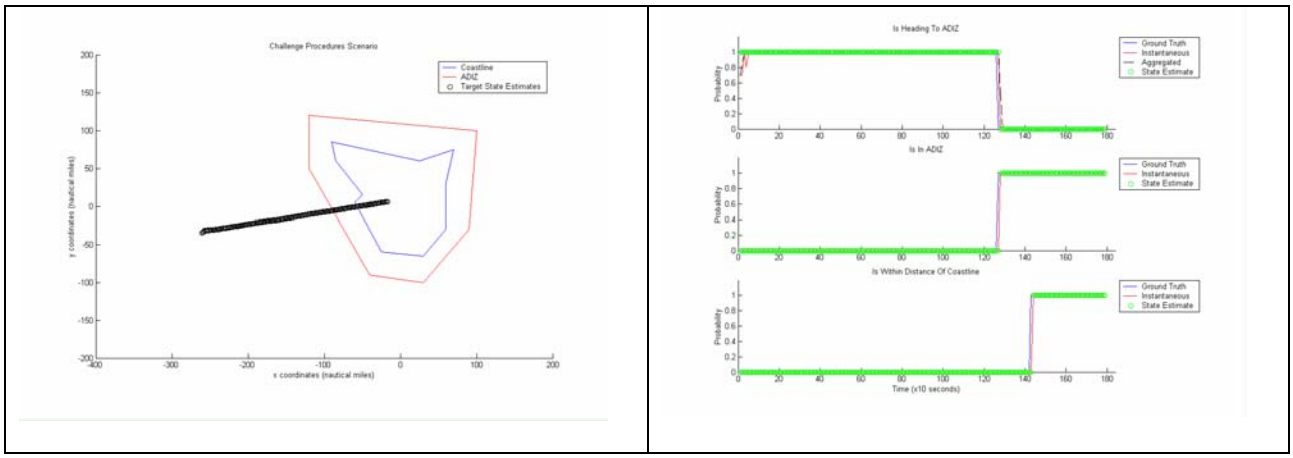


Figure 6: Target behaviour extraction methodology used to establish target-entity relations for monitoring a target's progress.

In summary, the methodology permits the behaviour of air targets to be interpreted automatically in a consistent and reasonably systematic manner that emulates the visual determination of the behaviours by air defence operators. It also has a number of desirable features such as it supports code re-use, and it provides a convenient representation of information which is flexible enough to support additional automated reasoning, whether it be by Bayesian networks, templates or intelligent software agents. Finally, the extracted information, summarised over time, may also be treated as "features" of a target track for the purposes of information retrieval and data mining (in particular clustering) to assist intelligence analysts.

4.0 INTELLIGENCE PROCESSING AND ANALYSIS

The process model employed by the intelligence community, aptly referred to as the *intelligence cycle*, comprises the five steps of *planning and direction*, *collection*, *processing*, *analysis and production*, and *dissemination*. It is evident from this model that, at a conceptual level, the business practices of the intelligence community are very similar to those employed for command and control in the defence forces. So too are the demands on intelligence analysts in regard to processing and analysing large volumes of data and information, although in the intelligence domain more emphasis is placed on knowledge discovery via data mining and information extraction.

Raw (unprocessed) intelligence data may be collected from many disparate sources via electronic and other technical means, human sources, imagery and communications for example. However, since most raw intelligence data takes the form of unstructured free text, the discussion in this paper is restricted (almost) exclusively to automated document analysis, although some of the issues raised are relevant to all forms of intelligence.

Just as the automatic processing of sensor data and the subsequent extraction of information from it is a challenging undertaking, so too is the automatic processing of documents for extracting intelligence information. However, unlike the analysis of sensor data, the automatic analysis of documents is not susceptible to treatment by numerical techniques. Thus, automated document analysis requires entirely different techniques from sensor data analysis, and must contend with the notorious difficulties that dealing with unstructured free text poses.

The general framework for intelligence information management and analysis adopted by DSTO is illustrated by the schematic in Fig. 7. With respect to this framework, there are three main challenges to be faced in introducing automation to document analysis. The first challenge lies in efficiently retrieving the documents relevant to a query. As any user of search engines on the world wide web knows, it can be extremely difficult to retrieve all relevant documents during an information retrieval query, without also retrieving a sizeable quantity of irrelevant documents. However, even assuming that individual intelligence sources, typically databases, have perfect information retrieval capabilities, often there is no connectivity between the sources, which again hampers efficient information retrieval. This lack of networking can result in the same document being retrieved multiple times because each of the sources needs to be queried individually. The second challenge is related to the processing of the retrieved documents. Information retrieval can recover documents for human analysis, but further processing of the documents for applications such as data mining and information fusion requires the unstructured information in the documents to be transformed into a well-structured form. Information extraction is a key enabling technology that can create the structured information required for these applications. However, extraction of key facts embedded within a piece of free text can prove to be problematic for a variety of reasons. For example, related facts of interest in a document may not be juxtaposed, but may instead be separated by a string of words which are extraneous to the facts, making it difficult to establish their association. In addition, references to the same fact may appear within a single document in different guises. As a case in point, dates referring to a common event may appear in either numerical or textual

form, and may be written in full, abbreviated or referred to descriptively, for example 11/9/01, 11 September 2001, 11 Sept. '01, 9/11, and "the day that the twin towers collapsed". Different conventions for representing dates such as the month-day-year format preferred in the United States and the day-month-year format used in Australia may also lead to ambiguity. The third challenge is to exploit the extracted information for the purposes of data mining, either by means of unsupervised learning (clustering), link analysis (association discovery and sequence discovery) or information visualisation, and ultimately for situation and impact refinement, including the recognition of indications and warnings.

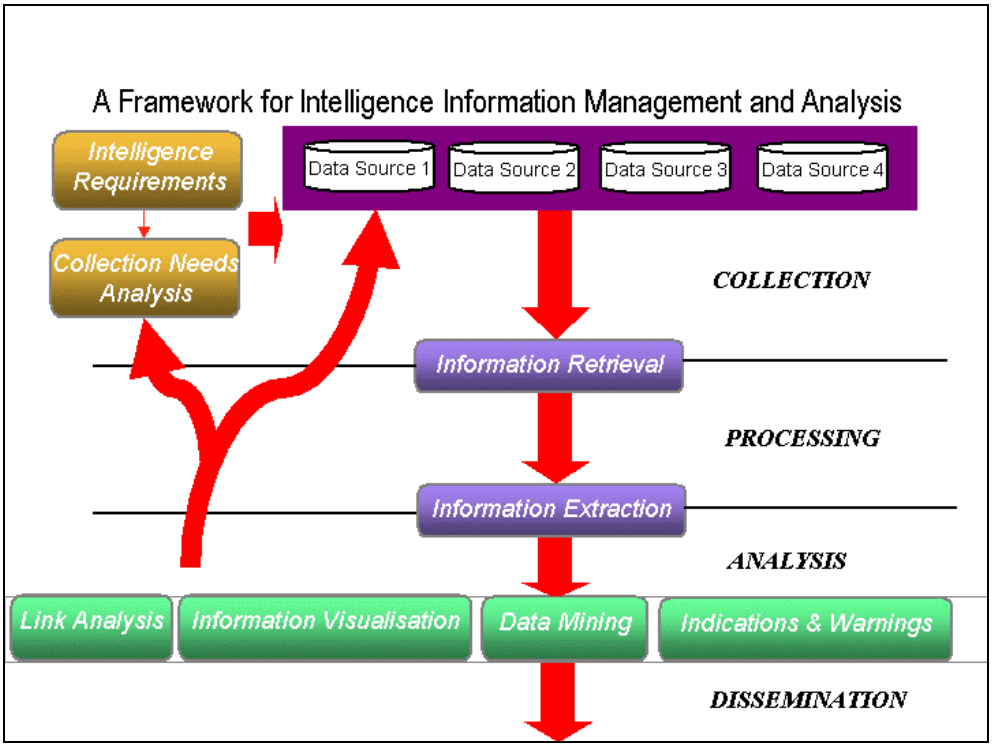


Figure 7: Schematic for intelligence information management and analysis [19].

DSTO's strategy for supporting automated text-based intelligence processing and analysis is based around these three challenges. It is evaluating a number of leading text processing and data mining commercial-off-the-shelf (COTS) products, as well as developing several in-house data mining toolkits which may assist in meeting the challenges. In the remainder of the section, a selection of these products will be profiled in brief, with particular emphasis placed on the in-house toolkits.

4.1 COTS Products[6]

Except where otherwise indicated, information on the three products below has been sourced from the internet websites of their respective manufacturers.

Autonomy™ is a product of Global Linxs that provides a software framework for automating operations on unstructured information. Some of the operations it supports include automatic categorisation, clustering, profiling, personalisation, alerting, hyperlinking and retrieval. Autonomy™ solutions are built using a modular architecture. In particular, Autonomy™ has a "portal-in-a-box" module which comprises three portals each with its own "portlets" that contain applications for performing a variety of functions.

[6] Disclaimer: The inclusion of these particular products in this paper is purely to illustrate the diversity of commercially available text processing and data mining functionality, and does not constitute their endorsement by DSTO.

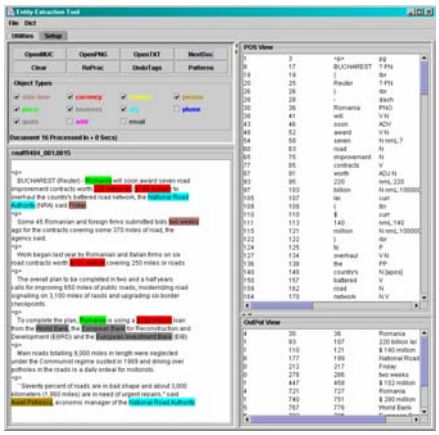
The *content portal* contains an information retrieval portlet which, like some search engines, supports concept-based retrieval and automatic summarisation of a document's content for example. However, it also features *classification* and *agent* portals. The classification portal identifies concepts in a document's contents and builds hierarchical classification schemes or categories. The agent portal performs personalisation operations using agents as concept patterns; in particular, it can use concept patterns to produce profiles to personalise information retrieval for users, recognise and alert users to highly focussed experts, and form communities of interest. For further details, refer to [20, 21].

SaffronNet™ is a product of Saffron Technology that provides the ability to process and learn about complex data in distributed multi-agent environments. It functions by establishing and maintaining a dynamic network of entities similar to a semantic network that automatically defines the context-based relationships or links between entities in the network. For further details, refer to [22].

Starlight™ is a product of Pacific Northwest National Laboratory, U.S. Department of Energy that comprises a diverse range of interactive data visualisation tools that allow the user to graphically manipulate the information under scrutiny and to display it in a number of different views. The Starlight™ information graphics fall broadly into two classes; the *non-spatial information graphics* provide spatial representations of non-spatial information such as text and numeric data, while the *inherently spatial information graphics* provide depictions of spatial information such as geospatial and CAD data. For further details, refer to [23].

4.2 In-House Products

The Fact Extractor System (FES) [24] is an information extraction product that is designed to extract "facts" from unstructured free text based on regular expression constructs. It functions on two levels. On the first level, which is commonly known as *named entity* extraction, it extracts facts of the form "who", "what", "where", and "when". On the second level, it functions by linking the basic facts extracted from the first level process into *events of interest*. The FES outputs three types of information: facts, event notifications and marked-up documents. The FES consists of a number of components; these are the fact extraction engine, the named entity extractor, a workbench for developing and testing fact extractor specifications, a batch tool for doing unsupervised information extraction, event alerting and document mark ups, and a formfiller for doing more advanced supervised information extraction. Figure 8 depicts a marked up document that has been outputted by the named entity extractor. The different entity types are distinguished by colour.



BUCHAREST (Reuter) - Romania will soon award seven road improvement contracts worth 220 billion lei ($140 million) to overhaul the country's battered road network, the National Road Authority (NRA) said Friday.
<p>
Some 45 Romanian and foreign firms submitted bids two weeks ago for the contracts covering some 370 miles of road, the agency said.
<p>
Work began last year by Romanian and Italian firms on six road contracts worth $102 million covering 250 miles or roads.
<p>
The overall plan to be completed in two and a half years calls for improving 650 miles of public roads, modernizing road signalling on 3,100 miles of raods and upgrading six border checkpoints.
<p>
To complete the plan, Romania is using a $280 million loan from the World Bank, the European Bank for Reconstruction and Development (EBRD) and the European Investment Bank (EIB).

Figure 8: Output of the named entity extractor in the form of a marked up document [24].

Finally, each set of fact extraction specifications determines a different fact extractor, so since each fact extractor functions as a software agent, different fact extractors can work cooperatively to reduce complexity for advanced information extraction. For more information, refer to [24].

The Data Mining and Visualisation Toolkit (DMVT)[7] [25] is a suite of prototype tools designed to assist in the analysis of historical positional and kinematic data produced through the tracking of dynamic entities in some problem domain, such as air targets in a wide area surveillance scenario for example. The toolkit is composed of five applications that are integrated as layers in the open source OpenMap Java toolkit from BBN Technologies [26]. Some of the strengths of the DMVT, which supports the analysis of temporal, spatial, link and track information, are its abilities to filter data for human-assisted data exploration and automatically determine normalcy patterns from the data. The DMVT features a number of customised views. For example, temporal data based on a particular activity may be displayed in a variety of modes on a grid or clock face based on "hour of the day", "day of the week" et cetera through to "month of the year" according to the user's needs, with the level of intensity of the activity for each time unit indicated via colour-coding. Similarly, target movements between given locations on a geospatial display may be indicated by colour-coded arrows, such that different colours indicate different classes of tracks and the width of the arrow indicates the volume of traffic between the locations.

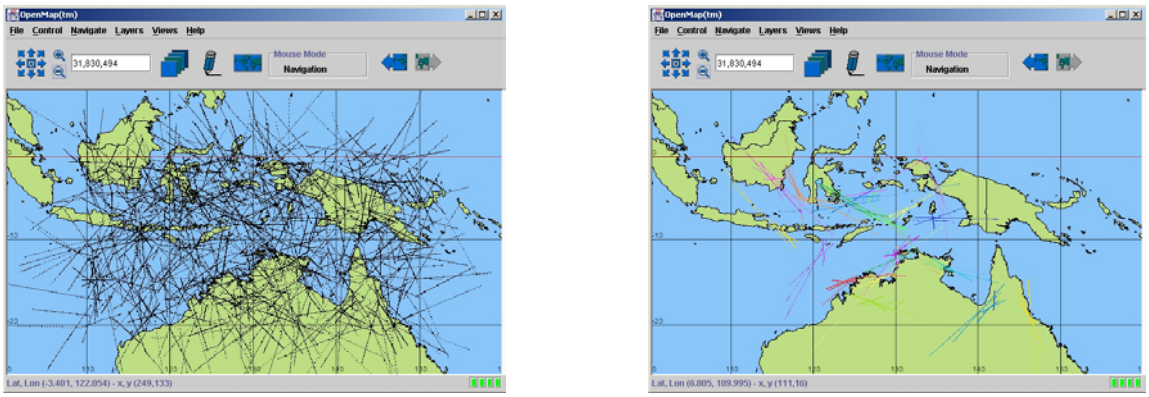


Figure 9: A set of synthetic track data before (left) and after track clustering and filtering (right) [25].

Finally, by combining the spatial view and the track clustering functionality, a set of target tracks may be clustered into classes of tracks with similar headings and close spatial separation and then viewed on the geospatial display. This technique is illustrated in Fig. 9 for a randomly generated set of synthetic air tracks (NB: no sensor model has been employed in their generation). The tracks have been clustered and clusters with a lone track have been removed to reveal potential air corridors. For more information, refer to [25].

5.0 INFORMATION FUSION FOR COMMAND & CONTROL

In Sections 3 and 4, automated data fusion and data mining systems were described for assisting defence operators and intelligence analysts in the processing and analysis of data and information. However, the important issue of human interaction with the systems was not raised. In this section, human computer interaction in information fusion is addressed in terms of DSTO's Future Operations Centre Analysis Laboratory (FOCAL) programme (FOCAL itself appears in Fig. 10). Before discussing the research being undertaken for FOCAL, it is necessary to probe more deeply into the interpretations of the λJDL model of

[7] The DMVT has been included, even though it is not a text-based information extraction product, because of its relevance to the material in Section 3.

data fusion to establish the context for the research. Note: The material for this section draws substantially from reference [12].

5.1 Interpretations of the λJDL Model of Data Fusion

The FOCAL data fusion system extends well beyond the traditional "machine sensor fusion" emphasis of the data fusion community, by including information fusion considerations involving both humans and machines. It was these considerations that led to Lambert's deconstruction of the US JDL model of data fusion in the form of the λJDL model described in Section 2. As is explained below, the λJDL model not only provides a means of integrating data fusion and data mining, but is also crucial for developing the theory which underpins the research within the FOCAL programme.

Figure 10: The Future Operations Centre Analysis Laboratory (FOCAL).

As a consequence of the FOCAL data fusion system seeking to fuse processes involving both people and machines, three distinct types of processes result: psychological processes associated with people; technological processes characteristic of machines; and integration processes facilitating interaction between the psychological and technological processes. Different interpretations of the λJDL model are obtained when different combinations of these processes are considered. Figure 11 exhibits a matrix in which the rows are the three components of the λJDL model and the columns are the three types of processes.

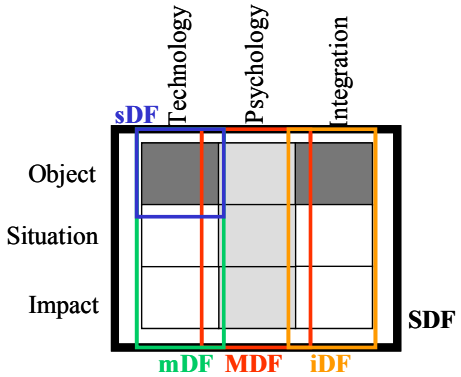


Figure 11: Interpretations of the λJDL model [12].

Five interpretations of the λJDL model are overlaid on the matrix. These interpretations are as follows:

Mental Data Fusion: Endsley [27] defines situation awareness as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". In [28], Lambert indicates that if "representation" in the definitions of the levels of the λJDL model is interpreted as "mental representation", then the object, situation and impact fusion components of *mental data fusion* (MDF) correspond to the situation awareness elements of perception, comprehension and projection respectively. In this way, mental data fusion and situation awareness may be identified.

Sensor Data Fusion: *Sensor data fusion* (sDF) refers to the technological interpretation of object refinement and corresponds to the commonly accepted definition of sensor fusion.

Machine Data Fusion: *Machine data fusion* (mDF) arises if "representation" in the definitions of the levels of the λJDL model is interpreted as "machine representation". It is concerned with technological approaches to data fusion and so refers to the technology column of the matrix in Fig. 11. Clearly, sensor data fusion is part of machine data fusion.

Interface Data Fusion: *Interface data fusion* (iDF) arises if "representation" in the definitions of the levels of the λJDL model is interpreted as "interface representation". It is concerned with human-machine integration approaches to data fusion, and so refers to the integration column of the matrix in Fig. 11.

System Data Fusion: *System data fusion* (SDF) is concerned with data fusion systems involving machines, people and interactions between the two. System data fusion refers to the entire matrix in Fig. 11. It requires a unified framework across all three components of the λJDL for handling interactions between people, interactions between machines, and interactions between people and machines.

These five interpretations of the λJDL model lead to a number of challenges for information fusion. This is expanded on in the sequel.

5.2 FOCAL Research – The Grand Challenges of Information Fusion

None of the elements of the matrix in Fig. 11 represents solved problems, and so all elements of the matrix pose challenges. However, the portions of the matrix which are unshaded pose grander challenges than the shaded portions. It is these *grand challenges*, which may be broadly categorised as either machine data fusion or system data fusion challenges, that form the basis for the research under the FOCAL programme. In [12], Lambert gives a full account of these challenges, but that level of detail is beyond the scope of this paper. Therefore, in the remainder of the section, only a brief outline of these grand challenges and a description of the related FOCAL research initiatives for each are given. For further details, refer to [12].

The Semantic Challenge: What symbols should be used and how do these symbols acquire meaning?

The semantic challenge transcends philosophical, mathematical and computational dimensions. A philosophical theory is required to conceptualise the domain of interest; a mathematical theory is required to impose structure on that conceptualisation; and a computational theory is required to bring that conceptualised structure to life. The conceptualisation suggests symbols that might be used. The (formal) mathematical theory prescribes the meaning of those symbols.

To accomplish this within the FOCAL programme, the intention is to develop formal theories to capture the semantics of selected primitive symbols [29]. The FOCAL programme needs to address a range of issues spanning multiple levels. Table 3 specifies the proposed levels, together with *some* relation

symbols to be defined for each level. In developing the formal theories, the ambition is not to capture *the* meaning of time, belief, et cetera, but to formulate *a* meaning of time, belief et cetera that is sufficient for engineering data fusion systems. Ideally, the metamathematical properties of the theories will include *soundness*, *completeness* and *decidability*, but these properties are often difficult to obtain.

Table 3: Proposed levels and some relation symbols for the formal theories [12].

Social:	group, ally, enemy, neutral, own, possess, invite, offer, accept, authorise, allow
Intentional:	individual, routine, learnt, achieve, perform, succeed, fail, intend, desire, belief, expect, anticipate, sense, inform, effect, approve, disapprove, prefer
Functional:	sense, move, strike, attach, inform, operational, disrupt, neutralise, destroy
Physical:	land, sea, air, outer_space, incline, decline, number, temperature, weight, energy
Metaphysical:	exist, fragment, identity, time, before, space, connect, distance, area, volume, angle

The Epistemic Challenge: What information should be represented, and how should it be represented and processed within the machine?

While the semantic challenge for information fusion relates to the choice of symbols and what those symbols mean, the epistemic challenge involves the choice of knowledge content and the choice of a symbolic knowledge representation scheme. For analytically difficult domains in which human expertise is available, as is often the case for information fusion applications, one approach toward the epistemic challenge is to model the cognitive behaviour of people. This requires a modelling framework, a means of capturing cognitive behaviour within that framework, and a means of automating that captured cognitive behaviour within a machine. This is the approach being taken for FOCAL.

In [30], an approach for automating cognitive routines for FOCAL is outlined, based on people's explanations of what they do, *as they are doing it*, to mitigate *a posteriori* rational reconstructions. In its current form, the explanations are recorded through speech recognition before a human translation process converts them into cognitive routines executable in the ATTITUDE multi-agent reasoning system. The ATTITUDE system [31], developed by DSTO, is so named because it codes in terms of *propositional attitudes* (beliefs, desires, expectations, et cetera), which significantly eases the burden of translation from explanation to code.

The Paradigm Challenge: How should the interdependency between the sensor fusion and information fusion paradigms be managed?

The lack of an established technological information fusion approach follows from the fact that the techniques of sensor fusion do not easily scale up. As noted in [28, 32-34], there is a paradigm shift in moving from sensor fusion to information fusion, a shift from an Aristotelian world of objects with measurable properties to a Wittgensteinian world of symbolically expressed facts formed from relations between objects. The technological aspects of sensor fusion are founded on an Aristotelian paradigm based on a numerical representation, while the technological aspects of information fusion are founded on a Wittgensteinian paradigm based on a symbolic representation. The paradigm challenge arises because machine data fusion demands an interdependency between these two paradigms, and perhaps in time, a unifying paradigm.

It has become customary to think of "levels" of data fusion, despite the bus architecture in the US JDL model in Fig. 2. The notion of levels follows from a presumption that the output of sensor fusion becomes

the input of information fusion. While not all inputs to information fusion derive from conventional sensor fusion, a dependency of information fusion on sensor fusion is widely recognised.

The FOCAL programme has not yet addressed the paradigm challenge, although it does support a hybrid numerical-symbolic approach, including a probabilistic inference engine within ATTITUDE [35]. Arguably, the paradigm challenge is best handled in concert with the communities processing the sensor data to ensure that the transformation from numerical sensor data to symbolic information does not lead to misinterpretations of the data. The information extraction efforts at DSTO for wide area surveillance and intelligence processing and analysis, reported on in Sections 3 and 4, support this viewpoint. It is envisioned that the information extracted from target track data and text-based information (and indeed other data sources) will be structured according to the social, physical and metaphysical formal theories arising from the semantic challenge.

The Interface Challenge: How should people be interfaced to complex symbolic information stored within machines?

As symbolic information fusion matures within the machine, a means is required of interfacing complex information encapsulated in symbolic machine representations to people interacting with those machines.

The FOCAL programme is responding to the interface challenge through a number of initiatives. The first of these initiatives is the *Virtual Battlespace* which is shown in Fig. 12. It advances beyond the typical 2-dimensional displays, which overlay target symbology or tracks (and other contextual information) on a geographic plot, by incorporating terrain, imagery, and objects of interest within a 3-dimensional stereoscopic display.

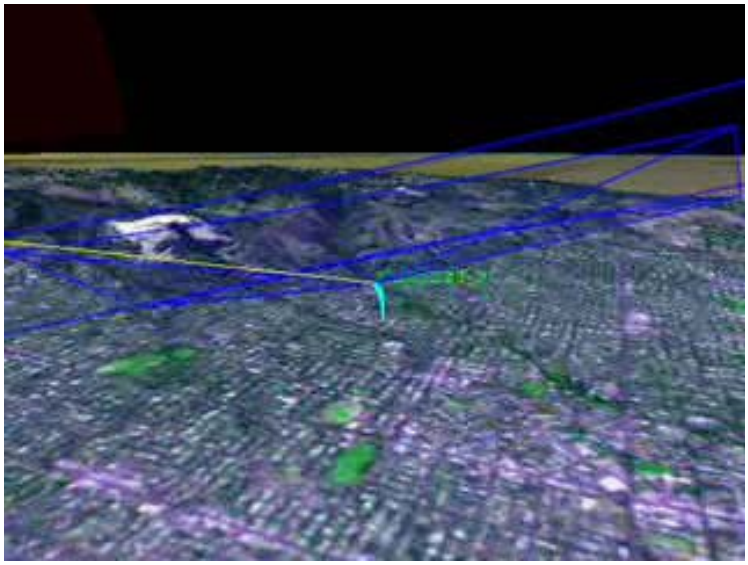


Figure 12: A snapshot of the 3-dimensional Virtual Battlespace [12].

However, while the Virtual Battlespace significantly advances the interface challenge for sensor fusion, it only marginally improves it for information fusion. The Virtual Battlespace is still not an appropriate medium for conveying information such as "Osama Bin Laden has avoided capture". In everyday life, information of this sort is customarily received via news services, be they print, radio, television or internet based. Television news broadcasts remain the most dominant form, and are typically composed of: (a) *advisers*, such as news presenters, weather presenters, sports presenters, reporters, experts; (b) *maps and diagrams*, such as weather maps and finance charts and graphs; and (c) *video footage*, which can be

live, or extracted from a library because of some relevance to news events. In [31], it is proposed that software counterparts be developed for these three components of news broadcasts. As software, the components provide *portability* and *interactivity*. Portability arises because the software can be readily distributed throughout a computer network, enabling people interacting with the system to access it on demand. Interactivity arises because, unlike the information push only mechanism of television news, the software can support an information push-pull interaction with people. This is the approach being taken in the FOCAL programme.

In FOCAL, the software counterparts of advisers, maps, diagrams and video footage are *Virtual Advisers*, the *Virtual Battlespace*, *Virtual Interaction* and *Virtual Video* respectively. The Virtual Advisers in FOCAL have two semblances, the cartoon-like Franco and the photo-realistic Jane who appear in Fig. 13. The visemes, phonemes and the emotional look of the characters, are dynamically specified by marked up text. This enables the characters to be dynamically controlled by people, scripts or the ATTITUDE software. Preliminary interactive speech dialogue with the characters has also been established.

Figure 13: FOCAL's Virtual Adviser Avatars – Franco on the left and Jane on the right [12].

The Virtual Battlespace has already been described. Virtual interaction involves the representation and visualisation of structured information, as well as the use of wands, hand gesturing and eye tracking. Finally, Virtual Video involves the automated reconstruction of 3-dimensional stereoscopic animations from formal knowledge representations of the sort alluded to in Table 3. Virtual Video combines the selection of animated models, the choice of camera perspectives, and the movie director's craft. Figure 14 illustrates a virtual video from FOCAL. Reference [36] reports some of the underlying technical issues.

Figure 14: Virtual Video in FOCAL [12].

The System Challenge: How should data fusion systems formed from combinations of people and machines be managed?

The remaining challenge – the system challenge – concerns the desire for a unified framework for system data fusion that can account for interactions between people, interactions between machines, and interactions between people and machines.

The construction of ATTITUDE agents to improve machine useability has had the side effect of producing a conceptual framework that can be applied to both people and machines. As a consequence, the response to the system challenge for FOCAL has been to treat the FOCAL data fusion system as an *agent society* composed of people and machines. The same conceptual constructs can then be applied to both people and machines. With respect to the semantic challenge, the relations of Table 3 can be applied to both people and machines. With respect to the epistemic challenge, the proposal for capturing cognitive routines is about establishing functionalist cognitive routines common to both people and machines. With respect to the interface challenge, the virtual advisers allow the machines to interface more like people.

The management of these agent societies is also under consideration within the FOCAL programme. Perugini and Lambert [37] have been investigating the organisation of agent societies for logistics operations. This has recently led to a proposed Provisional Agreement Protocol [38]. In time this is expected to develop into a basis for a system of axioms to constrain the nature of contractual interactions between combinations of human and machine agents.

6.0 CONCLUSION

The need for Australia to develop an information capability, as laid out in Australia's strategic defence policy, has been explained in terms of its strategic circumstances, and the role of the Defence Science and Technology Organisation in supporting the development of this emerging capability has been highlighted. The λJDL model of data fusion has been outlined and used to establish the theoretical underpinnings for integrating the key enabling technologies of data fusion and data mining required for the information capability. Finally, against the backdrop of this model, several of DSTO's information extraction and fusion initiatives in the areas of wide area surveillance, intelligence processing and analysis and command and control have been discussed, and the potential for synthesis of these initiatives via the grand challenges of information fusion has been articulated.

7.0 ACKNOWLEDGEMENTS

The authors would like to thank Mr. Peter den Hartog for his assistance in producing the graphics used in Figure 1.

REFERENCES

[1] Australian Federal Government (2000), *Defence 2000 – Our Future Defence Force*, The Commonwealth of Australia Defence White Paper.

[2] T.G. Feeman (2002), *Portraits of the Earth – A Mathematician Looks at Maps*, Mathematical World, Vol. 18, American Mathematical Society.

[3] Contributing Editors (2003), *SBS World Guide, 11th Edition*, Hardie Grant Publishing.

[4] Australian Federal Government (2003), *Australia's National Security – A Defence Update 2003*, The Commonwealth of Australia.

[5] D.L. Hall and J. Llinas (Eds.) (2001), *Handbook of Multisensor Data Fusion*, The Electrical Engineering and Applied Signal Processing Series, CRC Press.

[6] A.N. Steinberg and C.L. Bowman (2001), *Revisions to the JDL Data Fusion Model*, Chapter 2 of Hall and Llinas [5].

[7] F.E. White, Jr. (1988), *A Model for Data Fusion*, in the Proceedings of the 1st National Symposium on Sensor Fusion, Vol. 2.

[8] A.N. Steinberg, C.L. Bowman and F.E. White, Jr. (1998), *Revisions to the JDL Data Fusion Model*, Joint NATO/IRIS Conference, Quebec City, Quebec, 19-23 October, 1998.

[9] E. Waltz (1998), *Information Understanding: Integrating Data Fusion and Data Mining Processes*, in the Proceedings of the 1998 IEEE International Symposium on Circuits and Systems "ISCAS '98", Vol. 6, pp. 553-556, Monterey, California, 31 May – 3 June, 1998.

[10] G. Piatetsky-Shapiro and W.J. Frawley (1991), *Knowledge Discovery in Databases*, AAAI Press / The MIT Press.

[11] U.M. Fayyad (1997), Editorial, *Data Mining and Knowledge Discovery*, Vol. 1, No. 1.

[12] D.A. Lambert (2003), *Grand Challenges of Information Fusion*, Proceedings of the Sixth International Symposium on Information Fusion "Fusion 2003", pp. 213-220, Cairns, Queensland, Australia, 8-11 July, 2003.

[13] M.G. Oxenham (2000), *Automatic Air Target to Airlane Association*, Proceedings of the International Symposium on Information Fusion "Fusion 2000", Paris, France, USA, 10-13 July, 2000.

[14] H-T. Ong, M.G. Oxenham and B. Ristic (2002), *Estimating Biases in Sensor Measurements using Airlane Information*, Proceedings of the Information, Decision and Control Conference "IDC '02", pp 187–192, Adelaide, South Australia, Australia, 11-13 February, 2002.

[15] H-T. Ong, B. Ristic and M.G. Oxenham (2002), *Sensor Registration Using Airlanes*, Proceedings of the International Symposium on Information Fusion, "Fusion 2002", Annapolis, Maryland, USA, 7-11 July, 2002.

[16] M.G. Oxenham (2003), *Using Contextual Information For Extracting Air Target Behaviour From Sensor Tracks*, Proceedings of the Signal Processing, Sensor Fusion, and Target Recognition XII Conference at the 17th SPIE Annual AeroSense Symposium, 21-23 April, 2003.

[17] H-T. Ong (2003), *Sensor Registration Using Airlanes: Maximum Likelihood Solution*, Proceedings of the SPIE Conference "Signal and Data Processing of Small Targets 2003", San Diego, CA, USA, 3-8 August, 2003.

[18] M.G. Oxenham (2003), *Enhancing Situation Awareness for Air Defence via Automated Threat Analysis*, Proceedings of the Sixth International Symposium on Information Fusion "Fusion 2003", pp. 1086-1093, Cairns, Queensland, Australia, 8-11 July, 2003.

[19] R. Price (ed.) (2003), *Proceedings of the DSTO Workshop on Information Retrieval, Processing and Analysis of Disparate Data Sources*, Canberra, Australia, 11 April, 2003.

[20] http://www.digibahn.com/pdfs/PB_Autonomy_Portal_in_a_Box_0402.pdf

[21] L. Alvino (2003), *An Overview of Autonomy*, Presentation in [19].

[22] http://www.saffrontech.com/pages/products/saffronnet.html

[23] http://starlight.pnl.gov/

[24] J. Das (2003), *A DSTO Prototype Information Extraction System*, Presentation in [19].

[25] B. Williams (2003), *Advanced Data Mining and Visualisation*, Presentation in [19].

[26] http://www.bbn.com

[27] M.R. Endsley (1995). Toward a Theory of Situation Awareness in Dynamic Systems, *Human Factors*, Vol. 37 No. 1 pp. 32 – 64.

[28] D.A. Lambert (2001), *Situations for Situation Awareness*, Proceedings of the Fourth International Conference on Information Fusion "Fusion 2001", Montreal, Quebec, Canada, 7-10 August, 2001.

[29] D.A. Lambert (2003), *A Computational Metaphysics: Part I Existence*, DSTO Research Report (in preparation), Information Systems Laboratory, Defence Science and Technology Organisation.

[30] D.A. Lambert (2003), *Automating Cognitive Routines*, Proceedings of the Sixth International Symposium on Information Fusion "Fusion 2003", pp. 986-993, Cairns, Queensland, Australia, 8-11 July, 2003.

[31] D.A. Lambert (1999), *Advisers with Attitude for Situation Awareness*, Proceedings of the 1999 Workshop on Defense Applications of Signal Processing, pp. 113-118, edited by A. Lindsay, B. Moran, J. Schroeder, M. Smith and L. White, La Salle, Illinois, USA.

[32] L. Reznik and V. Kreinovich (Eds.) (2003), *Soft Computing in Measurement and Information Acquisition*, Studies in Fuzziness and Soft Computing No. 127, Springer Verlag.

[33] D.A. Lambert (2003), *An Exegesis of Data Fusion*, Chapter 6 of Reznik and Kreinovich [32].

[34] D.A. Lambert (1999), *Assessing Situations*, in Proceedings of the 1999 Conference on Information, Decision and Control "IDC '99", pp. 503-508, IEEE Press.

[35] I. Fabian, and D.A. Lambert (1998), *First-Order Bayesian Reasoning*, in Advanced Topics in Artificial Intelligence, 11th Australian Joint Conference on Artificial Intelligence "AI '98", pp. 131-142, Springer-Verlag.

[36] M. Coleman (2003), *Technical Considerations when Selecting Commercial Applications on an SGI Onyx*, DSTO Technical Report (in preparation), Information Systems Laboratory, Defence Science and Technology Organisation.

[37] D. Perugini, D.A. Lambert, L. Sterling, A. Pearce (2002), *Agents for Military Logistics Planning*, 15th European Conference on Artificial Intelligence "ECAI 2002", Lyon, France.

[38] D. Perugini, D.A. Lambert, L. Sterling, A. Pearce (2003), *Distributed Information Fusion Agents*, Proceedings of the Sixth International Symposium on Information Fusion "Fusion 2003", pp. 86-93, Cairns, Queensland, Australia, 8-11 July, 2003.

DEPARTMENT OF DEFENCE
DEFENCE SCIENCE & TECHNOLOGY ORGANISATION
DSTO

Information Fusion and Extraction Priorities for Australia's Information Capability

Presented by: Neil Gordon
On behalf of: Martin Oxenham *et al*

NATO Military Data and Information Fusion Conference
Prague, Czech Republic

October 20-22 2003

DEFENCE
SCIENCE & TECHNOLOGY

Overview

- Australia's strategic circumstances and its need for an integrated information capability
- Outline a model for integrating data fusion and data mining
- Discuss several of the Defence Science and Technology Organisation's information extraction and fusion initiatives in the areas of:
  - Wide Area Surveillance
  - Intelligence Processing and Analysis
  - Command and Control

DEFENCE
SCIENCE & TECHNOLOGY

Australia's Strategic Policy on Defence

- Australia's Defence White Paper states:

  "Effective use of information is at the heart of Australia's defence capability. In part this is a reflection of a worldwide trend, as information technology is transforming the ways in which armed forces operate at every level. All forms of capability are being transformed by the innovative use of information technology. But this trend is more significant to Australia than to many other countries. Our strategic circumstances mean that innovative applications of different aspects of information technology offer Australia unique advantages."

DEFENCE
SCIENCE & TECHNOLOGY

Australia's Strategic Circumstances

Australia compared to the USA

Australia compared to Europe

- Australia is an island continent that ranks 6th in size on a worldwide scale
- However, it has a population of only 19.2 million
- Accordingly, it has only a small defence force (~50,000 permanent and ~25000 reserves)
- The region of Australia in most need of defence is the sea-air gap to its north, while most of its population resides on the eastern seaboard
- All these factors pose great challenges for the Australian Defence Force and Australia's intelligence agencies

DEFENCE
SCIENCE & TECHNOLOGY

Australia's Strategic Circumstances Ctd.

Australia compared to the USA

Australia compared to Europe

Current Defence Imperatives (include):

- Border protection against smuggling (of people, drugs and other contraband) and illegal fishing
- Homeland defence against the threat of:
  - Terrorist attacks by extremist groups
  - The use of weapons of mass destruction by rogue states
- Participation in coalition operations eg Peace-keeping efforts

DEFENCE
SCIENCE & TECHNOLOGY

Australia's Strategic Circumstances Ctd.

Australia compared to the USA

Australia compared to Europe

Specific goals of Australia's emerging information capability (include):

- Maintaining first-rate intelligence capabilities
- Developing a comprehensive surveillance system
- Developing an integrated command system

To achieve these goals requires innovative methods for automatically extracting and fusing information from voluminous, disparate and distributed data sources.

DEFENCE
SCIENCE & TECHNOLOGY

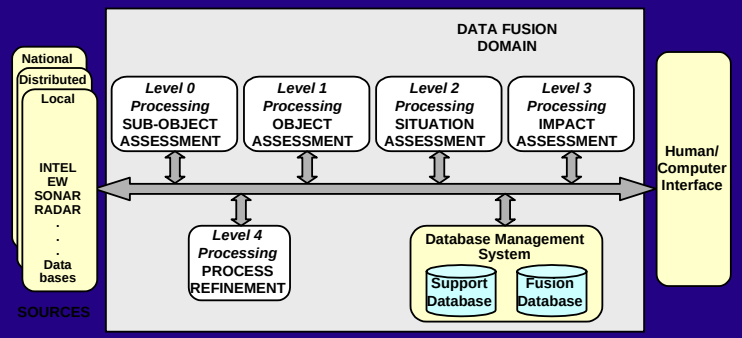
Key Enablers – Data Fusion & Data Mining

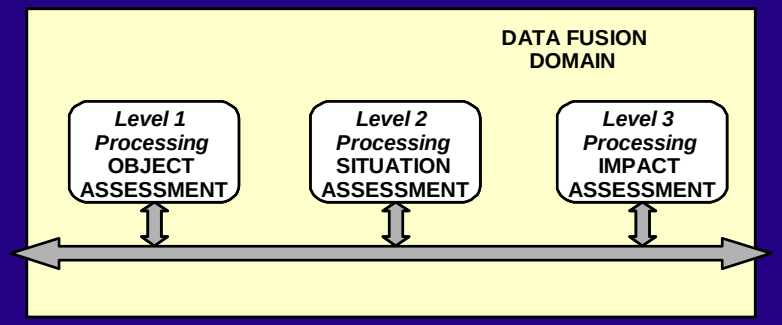
- *Data fusion* is commonly described as the process of combining data or information to estimate or predict entity states; while
- *Data mining*, which subsumes machine learning and data visualisation, may be regarded as the process of determining patterns, trends, relationships and associations in large data sets that are not explicitly mentioned in the raw data.
- Ideally, to fully exploit the available data and information, these processes should be *integrated*.

DEFENCE SCIENCE & TECHNOLOGY

Lambert's λJDL Model of Data Fusion

United States JDL Model





λJDL Model

- Lambert defines data fusion as "the process of utilising one or more data sources over time to assemble a representation of aspects of interest in an environment".
- He offers the following deconstruction of the well-known US JDL model of data fusion:
  - Level 1 – Identification of objects from their properties;
  - Level 2 – Identification of relations between these objects; and
  - Level 3 – Identification of the effects of these relationships between the objects.
- Under this model, data mining is subsumed as an integrated sub-process.

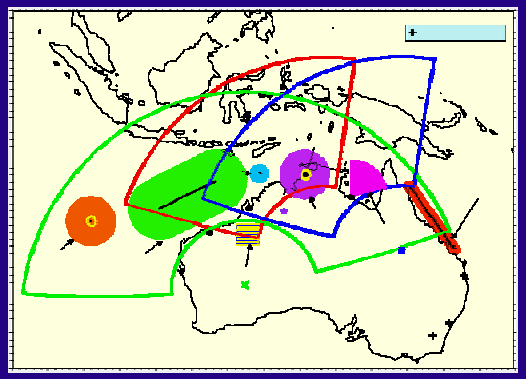
DEFENCE SCIENCE & TECHNOLOGY

Wide Area Surveillance – Target Behaviour Extraction

- Wide Area Surveillance (WAS) of the sea-air gap to the north of Australia plays a key role in Australia's information capability strategy.
- Data sources for WAS include:
  - Ground-based microwave radars (current);
  - Over-the-horizon radars (current);
  - Wedgetail Airborne Early Warning & Control capability (future);
  - Link 16 capability (future);
- The Wide Area Surveillance Picture, formed through the fusion of these data supports:
  - Real-time surveillance, fighter control, and situation & threat assessment (tactical level); and
  - Mission planning and intelligence gathering (theatre level).

DEFENCE
SCIENCE & TECHNOLOGY

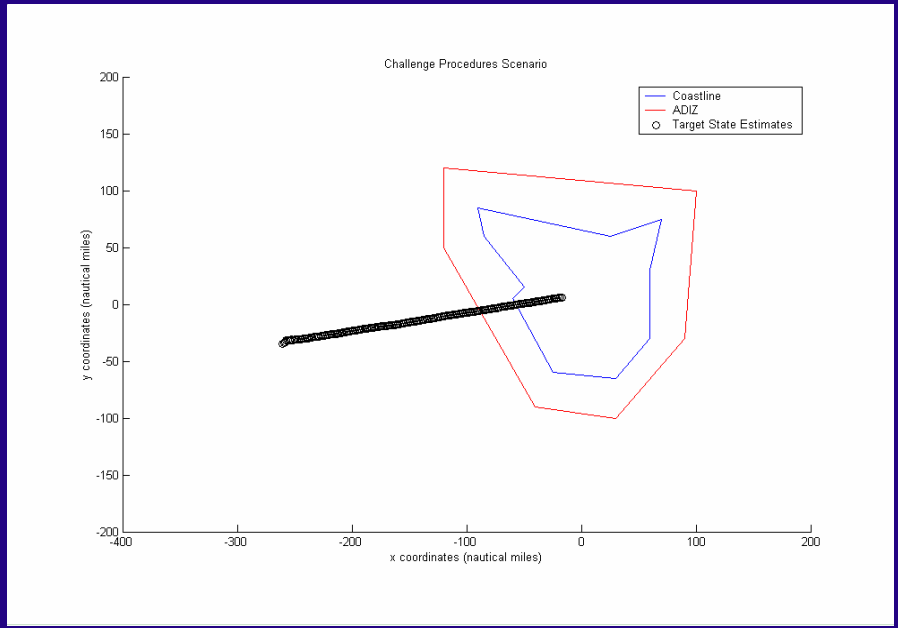
Target Behaviour Extraction



- Typically, compilation of the WAS picture and the activities it supports are performed manually.
- To allow operators to focus more on decision-making and less on data integration, it is desirable to automate some of the information processing required for these activities.
- This may be achieved in part by associating target tracks with the contextual information *ie entities* in the surveillance region to automatically extract the behaviours of the air targets.
- This contextual information may include for example:
  - Coastlines, airlanes & airports, shipping lanes & ports;
  - Locations of military assets & installations;
  - Regions eg Air Defence Identification Zones, Fighter Engagement Zones;
  - Notional radar coverage regions.

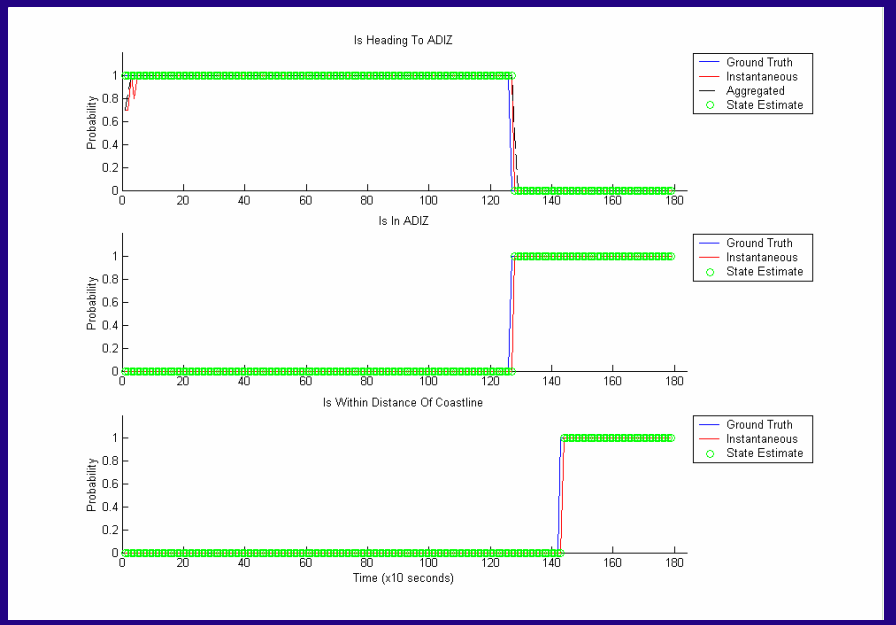
DEFENCE
SCIENCE & TECHNOLOGY

Example of Technique



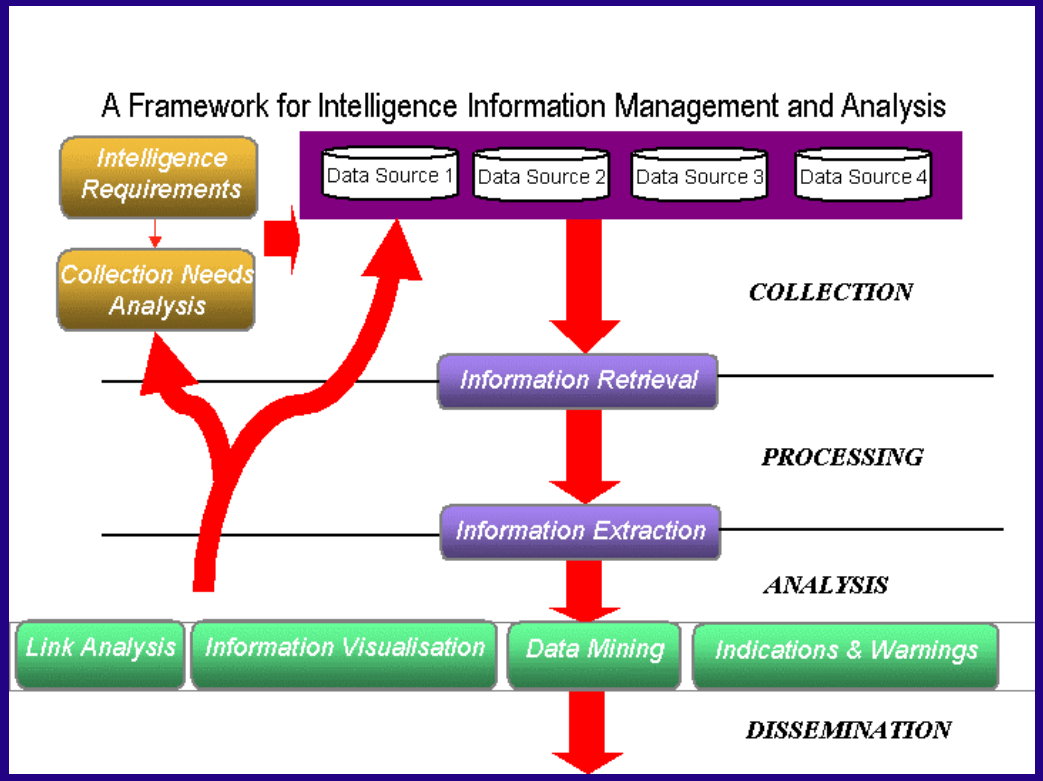
– The technique involves modelling the contextual information as simple geometric entities and establishing mathematical (geometric) criteria which hold if and only if the desired target behaviour is exhibited towards the entity.

– The resulting extracted information for each behaviour is represented as "Boolean-based" target-entity relations $\mathcal{R}(T,E)$.



DEFENCE
SCIENCE & TECHNOLOGY

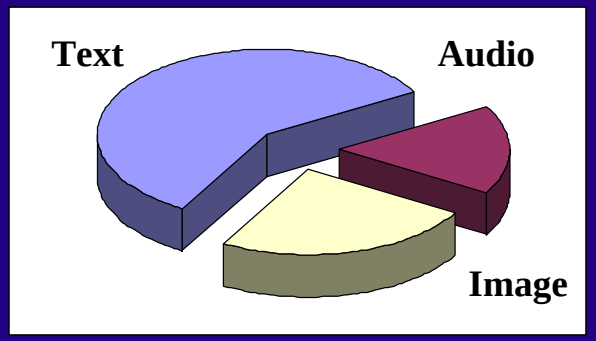
Intelligence Processing and Analysis



- DSTO's framework for investigating intelligence information management and analysis is aligned to the well-known *Intelligence Cycle*
  - Planning and direction
  - Collection
  - Processing
  - Analysis and production
  - Dissemination

DEFENCE
SCIENCE & TECHNOLOGY

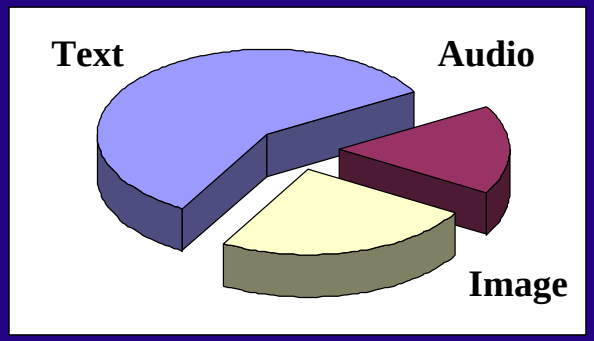
Three Main Challenges – Unstructured Free Text



- Efficient information retrieval
  - Lack of connectivity between databases
  - Duplication or irrelevance of retrieved documents
- Processing of unstructured information
  - Information extraction
    - Recognising key facts embedded in documents
    - Determining associations between related key facts
- Exploiting the extraction information
  - Data mining, link analysis, information visualisation and fusion

DEFENCE
SCIENCE & TECHNOLOGY

Data Mining and Information Retrieval Products

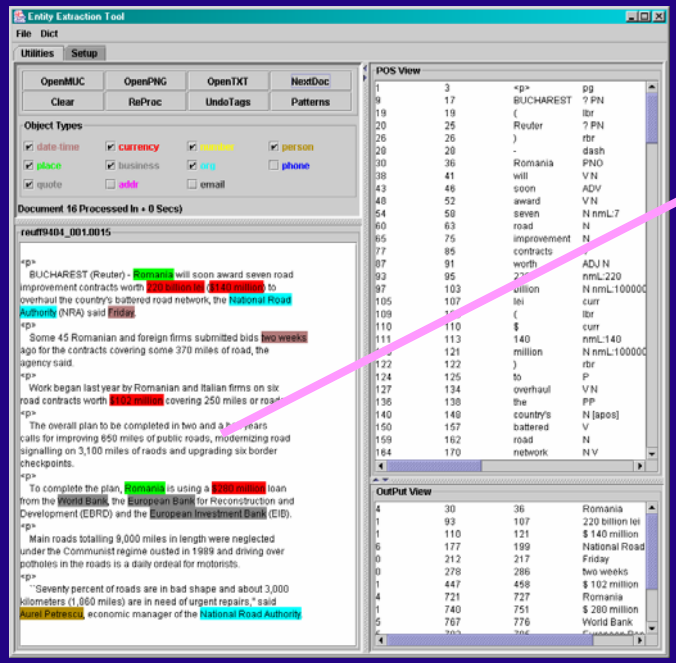


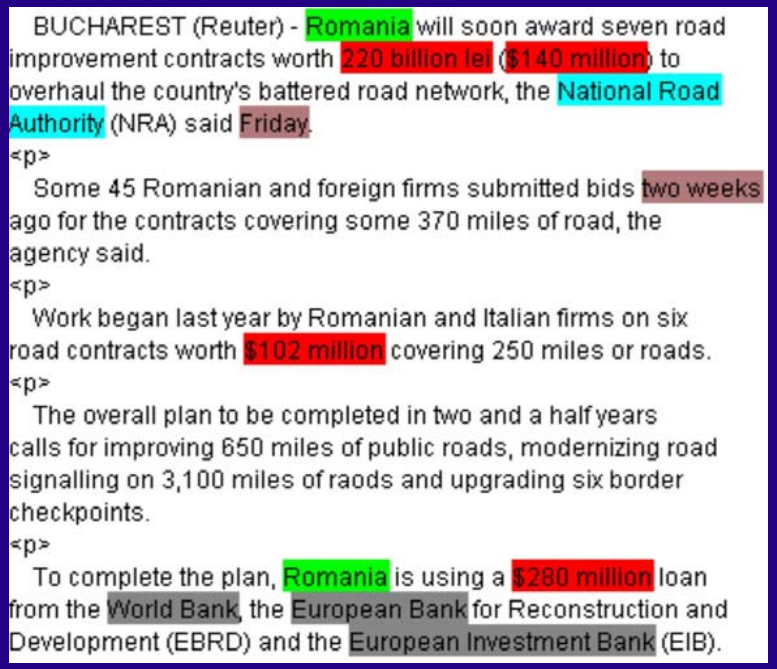
- COMMERCIAL PRODUCTS
  - Autonomy™
    - Distributed information retrieval, automatic concept extraction, generation of hierarchical classifications of documents etc.
  - SaffronNet™
    - Automatic determination of context-based relationships or links between entities in distributed multi-agent environments.
  - Starlight™
    - Advanced interactive data visualisation tools.
- IN-HOUSE PRODUCTS
  - The Fact Extractor System
  - The Data Mining and Visualisation Toolkit

DEFENCE
SCIENCE & TECHNOLOGY

The Fact Extractor System

- The Fact Extractor System works on two levels
  - It first extracts key facts from unstructured free text and then determines relations between the key facts using regular expression constructs
- It outputs facts, event notifications and marked-up documents (as illustrated below)
- Fact extractor agents may work cooperatively for advanced information extraction




BUCHAREST (Reuter) - Romania will soon award seven road improvement contracts worth 220 billion lei ($140 million) to overhaul the country's battered road network, the National Road Authority (NRA) said Friday.

<p>

Some 45 Romanian and foreign firms submitted bids two weeks ago for the contracts covering some 370 miles of road, the agency said.

<p>

Work began last year by Romanian and Italian firms on six road contracts worth $102 million covering 250 miles or roads.

<p>

The overall plan to be completed in two and a half years calls for improving 650 miles of public roads, modernizing road signalling on 3,100 miles of raods and upgrading six border checkpoints.

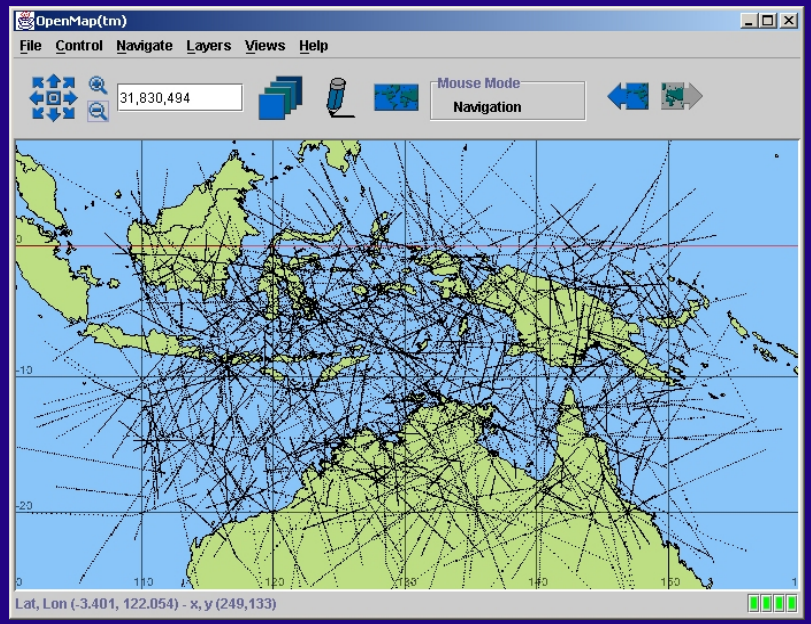
<p>

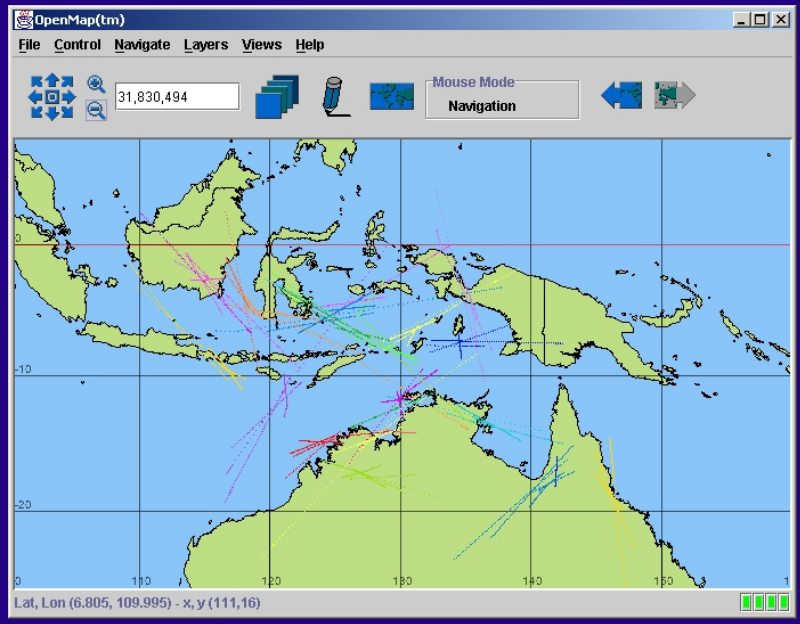
To complete the plan, Romania is using a $280 million loan from the World Bank, the European Bank for Reconstruction and Development (EBRD) and the European Investment Bank (EIB).

DEFENCE
SCIENCE & TECHNOLOGY

The Data Mining and Visualisation Toolkit

- Suite of prototype tools for analysing tracked entities based on their positional and kinematic data eg tracked land, sea and air targets
- Supports human-assisted data exploration of spatio-temporal, link and track information via clustering and filtering of the data
- Example: clustering and filtering performed on a randomly generated set of synthetic air target tracks (no sensor model employed)





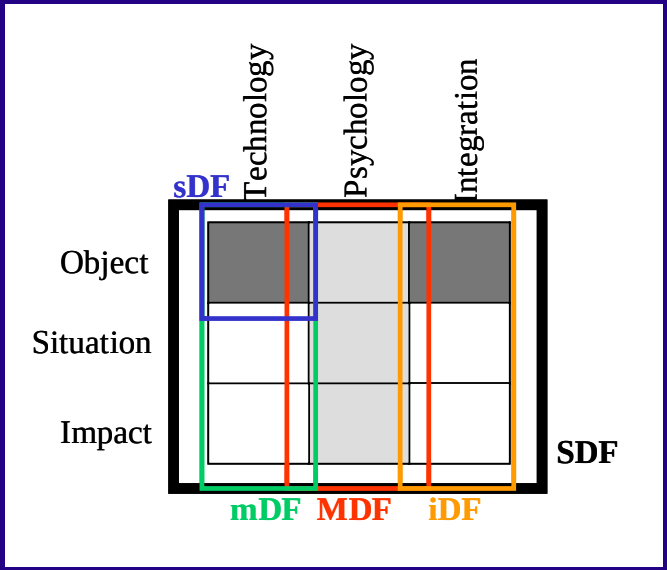
DEFENCE
SCIENCE & TECHNOLOGY

Information Fusion for Command and Control

- FOCAL – Future Operations Command Analysis Laboratory
  - Data fusion testbed for investigating information fusion processes involving both people and computers.
- ATTITUDE – Key enabling technology for FOCAL
  - Multi-agent reasoning system;
  - Developed by DSTO;
  - So named because it codes in terms of *propositional attitudes* eg beliefs, desires, expectations, etc.
- FOCAL employs the λJDL model of data fusion.

DEFENCE SCIENCE & TECHNOLOGY

Grand Challenges of Information Fusion



Interpretations of the λJDL Model

- Semantic Challenge – What symbols should be used and how do they acquire meaning?
- Epistemic Challenge – What information should be represented, and how should it be represented and processed within machines?
- Paradigm Challenge – How should the interdependency between the sensor fusion and information fusion paradigms be managed?
- Interface Challenge – How should people be interfaced to complex symbolic information stored within machines?
- System Challenge – How should data fusion systems formed from combinations of people and machines be managed?

DEFENCE SCIENCE & TECHNOLOGY

Grand Challenges of Info Fusion – FOCAL Research

FORMAL THEORIES ⟨Ω; +, •, -, ⊥, Ω⟩

Multiple Levels

Social: Group, ally, enemy, neutral, own, etc

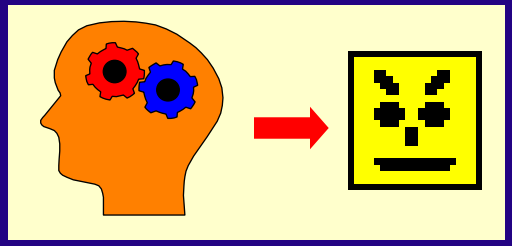
Intentional: Individual, routine, learnt, succeed, fail, etc

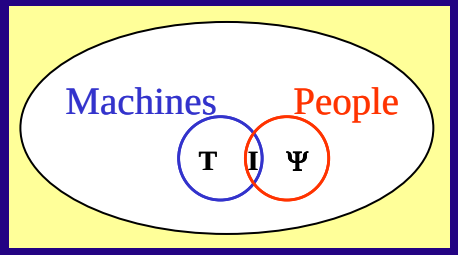
Functional: Sense, move, strike, attach, inform, etc

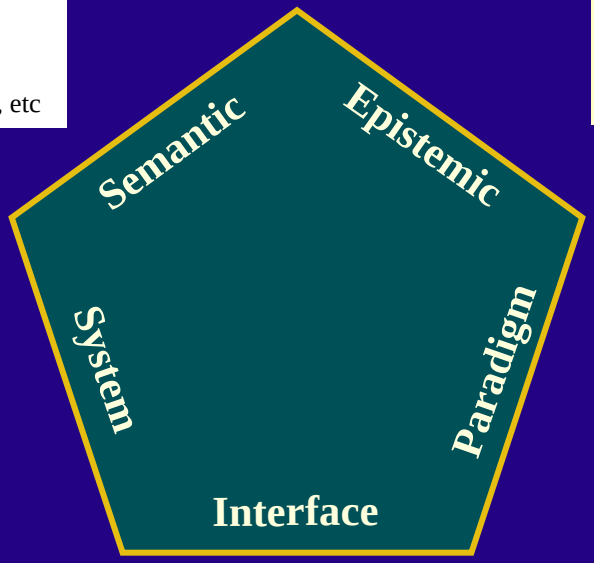
Physical: Land, sea, air, number, speed, weight, etc

Metaphysical: Exist, fragment, identity, time, space, etc

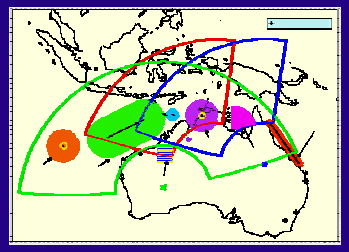
Automating Cognitive Routines

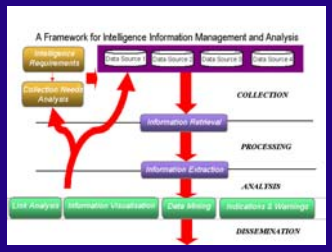


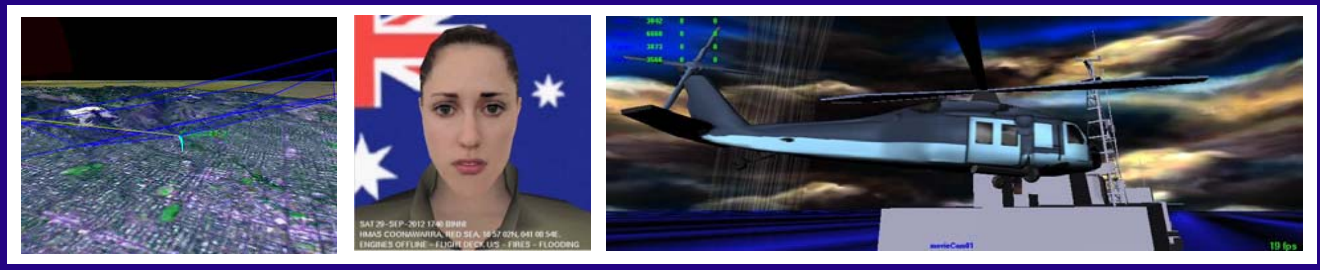




Information Extraction







Virtual News Broadcasts

DEFENCE
SCIENCE & TECHNOLOGY

Conclusion

- Highlighted Australia's strategic circumstances and its need for an integrated information capability
- Outlined an integrated model of data fusion and data mining in the form of the λJDL model
- Given an overview of several of the information extraction and fusion initiatives at the Defence Science and Technology Organisation in the areas of:
  - Wide Area Surveillance
  - Intelligence Processing and Analysis
  - Command and Control

DEFENCE
SCIENCE & TECHNOLOGY

Questions

?

martin.oxenham@dsto.defence.gov.au